STAT C206B: Topics in Stochastic Processes

Feynman Liang*
Department of Statistics, UC Berkeley

Last updated: January 30, 2020

Contents

1 Lecture 1: Backround material

2020-01-21

1.1 Ferguson distributions / Dirichlet processes

Definition 1 (*Gamma distribution*)

Random variable X supported on $(0, \infty)$ has *Gamma distribution* with shape $\alpha > 0$ and inverse scale / rate $\beta > 0$, written $X \sim \text{Gamma}(\alpha, \beta)$ if it has density

$$f_X(t) = \mathbb{1}\{t \in (0, \infty)\} \frac{\beta^\alpha t^{\alpha-1} e^{-\beta t}}{\Gamma(\alpha)} \tag{1}$$

where $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$ is the Gamma function defined for all $\Re t > 0$ and analytically continued to $\mathbb{C} \setminus \{n \in \mathbb{Z} : n < 0\}$

Proposition 2 (*Gamma closed under summation*)

{prop:gamma-closed-sum} *If* $Y \sim \text{Gamma}(\alpha, \beta)$ *and* $Z \sim \text{Gamma}(\gamma, \beta)$ *are independent, then* $Y + Z \sim \Gamma(\alpha + \gamma, \beta)$.

*feynman@berkeley.edu

Proof.

$$\begin{aligned}
f_{Y+Z}(t) &= \int_0^t f_Y(u) f_Z(t-u) du \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} \int_0^t u^{\alpha-1}(t-u)^{\gamma-1} du \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} \int_0^1 (tv)^{\alpha-1}(t-(tv))^{\gamma-1} t dv \\
&= \frac{1}{\Gamma(\alpha)\Gamma(\gamma)} \beta^{\alpha+\gamma} e^{-\beta t} t^{\alpha+\gamma-1} B(\alpha,\gamma)
\end{aligned}$$

where $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$ is the beta function □

A closely related distribution obtained from concatenating Gamma random variables into a vector and then normalizing the sum to 1 is the Dirichlet distribution.

Definition 3 (*Dirichlet distribution*)

Let $\boldsymbol{\alpha} \in (0,\infty)^K$. Random (probability) vector X taking values on the $K-1$-dimensional probability simplex $\Delta^{K-1} = \{\boldsymbol{x} \in [0,1]^K : \sum_i x_i = 1\}$ has *Dirichlet distribution* of order K and concentration parameters $\boldsymbol{\alpha}$, denoted $X \sim \mathrm{Dir}(\boldsymbol{\alpha})$, if it has density

$$f_X(\boldsymbol{x}) = \mathbb{1}\{\boldsymbol{x} \in \Delta\} \underbrace{\frac{\Gamma\left(\sum_{i=1}^K \alpha_i\right)}{\prod_{i=1}^K \Gamma(\alpha_i)}}_{=:B(\boldsymbol{\alpha})^{-1}} \prod_{i=1}^K x_i^{\alpha_i - 1}$$

Proposition 4 (*Constructing Dirichlet from Gammas*)

{prop:dirichlet-from-gamma} *Let $X_1, \ldots, X_n$ be independent* $\mathrm{Gamma}(\alpha_i, \beta)$ *distributed,* $S_n = \sum_{i=1}^n X_i$. *Then* $(V_i)_i = (X_i/S_n)_i \sim \mathrm{Dir}(\boldsymbol{\alpha})$.

Proof. $S_n \sim \Gamma(\sum_i^n \alpha_i, \beta)$ by Proposition 2 and for $\boldsymbol{v} \in \Delta^{n-1}$, we have

$$\begin{aligned}
f_V(\boldsymbol{v}) &= \int_0^\infty f_X(sv_1, \ldots, sv_{n-1}, sv_n) f_{S_n}(s) ds \\
&= \int_0^\infty e^{-\sum_{i=1}^n sv_i} \left(\prod_{i=1}^n \frac{(sv_i)^{\alpha_i - 1}}{\Gamma(\alpha_i)} \right) \frac{s^{\sum_i^n \alpha_i - 1} e^{-s}}{\Gamma(\sum_i^n \alpha_i)} ds \\
&= \frac{1}{\prod_1^n \Gamma(\alpha_i)} \prod_{i=1}^n v_i^{\alpha_i - 1} \int_0^\infty e^{-s\cancelto{1}{\sum_1^n v_i}} s^{(\sum_1^n \alpha_i)-1} ds \\
&= \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_1^n \Gamma(\alpha_i)} \prod_{i=1}^n v_i^{\alpha_i - 1}
\end{aligned}$$

□

Similar to Proposition 2 (Gamma closed under summation), where adding two Gammas yielded another Gamma where the parameters were added, Dirichlet distributions enjoy a similar kind of closure: "clumping" coordinate axes together (described below) yields another Dirichlet distribution where the parameters of the clumped axes are summed together.

Proposition 5 (*Dirichlet clumping property*)

Suppose $X \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_n)$. *For any* $r \le n$, *let* $V_i = X_i$ *for* $i \in [r]$ *and let* $V_{r+1} = \sum_{j=r+1}^n X_j$. *Then* $V \sim \mathrm{Dir}(\alpha_1, \ldots, \alpha_r, \sum_{j=r+1}^n \alpha_j)$.

Proof. By induction, it suffices to show this for $r = n - 2$. Notice

$$\begin{aligned} f(v_1, \dots, v_r, s) &= B(\boldsymbol{\alpha})^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) \int \mathbb{1}\left\{x_{n-1} + x_n = s\right\} x_{n-1}^{\alpha_{n-1}-1} x_n^{\alpha_n - 1} dx_{n-1} dx_n \\ &= B(\boldsymbol{\alpha})^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) \int_0^s u^{\alpha_{n-1}-1} (s-u)^{\alpha_n - 1} du \\ &= B(\boldsymbol{\alpha})^{-1} \left(\prod_{i=1}^{n-1} v_i^{\alpha_i - 1} \right) s^{\alpha_{n-1} + \alpha_n - 1} B(\alpha_{n-1}, \alpha_n) \end{aligned}$$

Since $\frac{B(\alpha_{n-1},\alpha_n)}{B(\boldsymbol{\alpha})} = \frac{\Gamma(\sum_1^n \alpha_i)}{\Gamma(\alpha_{n-1}+\alpha_n) \prod_1^{n-2} \Gamma(\alpha_i)}$, we are done. □

Iterating this result over coordinate axes other than the last $n-r$, we see that "clumping together" entries in a Dirichlet random vector yields another Dirichlet random vector with parameters also "clumped together." Concretely, for any mapping $\phi : [n+1] \to [m+1]$ if $U_j = \sum_{\phi(i)=j} V_i$ then U has Dirichlet distribution with parameters $\gamma_j = \sum_{\phi(i)=j} \alpha_i$.

Generalizing this clumping property is the motivation for *Ferguson Distributions* ferguson1973 [Fer73].

Definition 6 (***Ferguson / Dirichlet process distribution***)

Let μ be a finite positive Borel measure on complete separable metric space E. A random probability measure μ^* on E (i.e. a stochastic process indexed by a σ-algebra on E) has *Ferguson distribution with parameter* μ if for every finite partition $(B_i)_{i \in [r]}$ of E the random vector

$$(\mu^*(B_i))_{i \in [r]} \sim \mathrm{Dir}(\mu(B_1), \dots, \mu(B_r))$$

Lemma 7 (***Preservation of Ferguson under pushforward***)

Let μ^* be Ferguson with parameter μ and $\phi : E \to F$ measurable. Then the pushforward $\mu^* \circ \phi^{-1}$ is a random probability measure on F that has Ferguson distribution with parameter $\mu \circ \phi^{-1}$.

Proof. For $(B_i)_{i \in [r]}$ a finite partition of F, $(\phi^{-1}(B_i))_i$ is a finite partition of E. Since μ^* is Ferguson

$$(\mu^*(\phi^{-1}(B_i)))_i \sim \mathrm{Dir}((\mu(\phi^{-1}(B_i)))_i)$$

Hence $\mu^* \circ \phi^{-1}$ is Ferguson with parameter $\mu \circ \phi^{-1}$. □

Next, we turn to an important class of a Ferguson distributions arising from generalizing the Pólya urn.

Definition 8 (***Polya sequence***)

A sequence $(X_n)_{n \in \mathbb{N}}$ with values in E is a *Polya sequence with parameter* μ if for all $B \subset E$.

$$\begin{aligned} \Pr[X_1 \in B] &= \mu(B)/\mu(E) \\ \Pr[X_{n+1} \in B \mid X_1, \dots, X_n] &= \mu_n(B)/\mu_n(E) \end{aligned}$$

where $\mu_n = \mu + \sum_{i=1}^n \delta_{X_i}$.

Remark 9. When E is finite (e.g. a set of colors for the balls), (X_n) represents the result of successive draws from an urn with initially $\mu(x)$ balls of color $x \in E$ and after each draw a ball of the same color as the one drawn is added back to give an urn with color distribution $\mu_{n+1}(x)$.

blackwell1973 [BM73] gives the following result connecting Pólya sequences and Ferguson distributions.

Theorem 10 (***Polya Urn Schemes***)

Let (X_n) be a Polya sequence with parameter μ. Then:

1. *$m_n = \mu_n/\mu_n(E)$ converges almost surely to a limiting discrete measure μ^**

2. *μ^* has Ferguson distribution with parameter μ*

3. Given μ^*, $(X_i)_{i\geq 1}$ are independent with distribution μ^*

Proof. First consider E finite. Let μ^* and $\{X_i\}$ be random variables whose joint distribution satisfies (2.) and (3.).

Let π_n be empirical distribution of $(X_i)_{i\in[n]}$. $X_i \overset{\text{iid}}{\sim} \mu^*$, so by SLLN $\pi_n \overset{as}{\to} \mu^*$ and since

$$m_n = \frac{\mu + n\pi_n}{\mu(E) + n} \tag{2}$$

(1.) follows.

To complete the proof, we show equality in distribution of $\{X_i\}$ with a Polyá-μ sequence. This amounts to showing

$$\Pr[A] = \prod_x \mu(x)^{[n(x)]}/\mu(E)^{[n]} \tag{3}$$ {{eq:polya-seq-meas}}

where $A = \{X_i = x_i\}_i \in \{0,1\}^n$ and $n(x) = \#\{i : x_i = x\}$, and the rising factorial $a^{[k]} = a(a+1)\cdots(a+k-1)$. By the tower rule and $\{X_i\}$ IID

$$\Pr[A] = \mathbb{E}\left[\Pr[A \mid \mu^*]\right] = \mathbb{E}\left[\prod_x \mu^*(x)^{n(x)}\right] \tag{4}$$

Since μ^* is Ferguson, viewing $E = \sqcup_{x\in E}\{x\}$ as a partition we have $(\mu^*(x))_{x\in E} \sim \mathrm{Dir}((\mu(x))_{x\in E})$ so the RHS is the $(n(x))_{x\in E}$ moment of the Dirichlet distribution, which is equal to

$$\mathbb{E}\left[\prod_x \mu^*(x)^{n(x)}\right] = \frac{\Gamma(\mu(E))}{\Gamma(\mu(E)+n)}\prod_x \frac{\Gamma(\mu(x)+n(x))}{\Gamma(\mu(x))} = \frac{1}{\mu(E)^{[n]}}\prod_x \mu(x)^{[n(x)]} \tag{5}$$ {{eq:dirichlet-moment}}

as required by Eq. (3).

General E follows from approximation argument. □

Notice that the Dirichlet moment comparison in Eq. (5) was the key step relating μ to μ^*.

We leave the discreteness part of (1.) as an exercise, noting that similar to how Dirichlets can be defined as a set of independent Gammas normalized by their sum (Proposition 4 (Constructing Dirichlet from Gammas)) we would expect the Dirichlet process / Ferguson random measures to be definable as a gamma process with independent "increments" divided by their sum.

Exercise 11. Prove every Ferguson random measure is discrete. (Hint: argue using moments).

Remark 12. If (X_i) is a Polya sequence, then it is a mixture of IID sequences (each drawn from μ^*) with mixture weights given by the Ferguson distribution on μ^*. Hence, (X_i) is exchangeable i.e. $(X_i) \overset{d}{=} (X_{\sigma(i)})$ This is already apparent in Eq. (3), and more generally de Finetti's theorem guarantees that *any* exchangeable sequence is a mixture of IID sequences.

1.2 Construction of Haar Measure

For a finite group G, the measure $\mu(g) = \frac{1}{\#G}$ is left and right translation invariant i.e. $\mu(gA) = \mu(A) = \mu(Ag)$ for all $A \subset G$.. As we will prove, all compact groups have unique translation invariant measure, called the Haar measure.

Example 13

Let $Z_{ij} \overset{\text{iid}}{\sim} N(0,1)$ for $i,j \in [n]$ and X the Gram-Schmidt orthonormalization of the rows of Z. By rotation invariance of Z, we can show $XU \overset{d}{=} UX$ for all $U \in O(n)$, so X has Haar measure on the compact (Lie) group $O(n)$.

Definition 14

A *topological vector space* (TVS) is a vector space equipped with a topology such that vector space operations are jointly continuous.

Example 15

$\mathbb{R}^n$ with standard topology, any Banach space.

Definition 16

{def:equicontinuous}

A family $\mathfrak{G}$ of linear transformations on TVS $\mathfrak{X}$ is *(uniformly) equicontinuous on subset* $K \subset \mathfrak{X}$ if for every neighborhood V of the origin, there exists a neighborhood U of the origin such that

$$\forall k_1, k_2 \in K : k_1 - k_2 \in U \Rightarrow \mathfrak{G}(k_1 - k_2) \subset V$$

That is, $T(k_1 - k_2) \in V$ for all $T \in \mathfrak{G}$.

Remark 17. Whereas "uniform" is used in analysis to generalize the U neighborhood of continuity (e.g. the δ in ε-δ definition of continuity) from at a particular $x_0 \in \mathfrak{X}$ to $\forall x \in \mathfrak{X}$, "equi" is used to generalize from a single $f \in \mathcal{C}(\mathfrak{X})$ to a family $\mathfrak{G} \subset \mathcal{C}(\mathfrak{X})$.

Definition 18

A *locally convex topological vector space* (LCTVS) is a TVS with a local base of absolutely convex absorbing sets at the origin.

Definition 19 (*In-Class*)

A *locally convex topological vector space* (LCTVS) is a TVS such that the topology has a base consisting of convex sets.

To construct Haar measure for any compact group, we will need a fix point theorem due to Kakutani.

Theorem 20 (*Kakutani Fix Point Theorem*)

{thm:kakutani}

K compact convex subset of LCTVS $\mathfrak{X}$, $\mathfrak{G}$ group of linear transforms equicontinuous on K and such that $\mathfrak{G}(K) \subset K$, then $\exists p \in K$ such that

$$\mathfrak{G}(p) = \{p\} \tag{6}$$

Proof. By Zorn's lemma (consider inclusion chains of sets satisfying these properties, exploiting stability under uncountable intersection), there is some minimal compact convex $K_1 \subset K$ such that $K_1 \neq \emptyset$ and $\mathfrak{G}(K_1) \subset K_1$.

Since we are done if $K_1 = \{p\}$ is a singleton, assume otherwise. We will contradict minimality of K_1 by constructing $K_2 \subsetneq K_1$ such that $K_2 \neq \emptyset$ and $\mathfrak{G}K_2 \subset K_2$.

We first exploit equicontinuity to construct a minimal convex cover U of $K_1 - K_1$ such that $T(aU) = U$ for $T \in \mathfrak{G}$ and $a \in \mathbb{R}$ as well as $(1-\varepsilon)\bar{U} \not\supset K_1 - K_1$.

- By assumption, $K_1 - K_1$ (Minkowski sum) contains a point other than the origin, so there exists a neighborhood $V \ni 0$ such that $\bar{V} \not\supset K_1 - K_1$.
- Again using $\mathfrak{X}$ LCTVS, for some $|\alpha| \leq 1$, there is a convex neighborhood $V_1 \ni 0$ such that $\alpha V_1 \subset V$.
- By equicontinuity of $\mathfrak{G}$ on $K \supset K_1$, there is $U_1 \in N(0)$ such that for $k_1, k_2 \in K$ and $k_1 - k_2 \in U_1$ we have $\mathfrak{G}(k_1 - k_2) \subset V_1$.
- Let
$$U_2 := \operatorname{conv}(\mathfrak{G}U_1 \cap (K_1 - K_1)) = \operatorname{conv}(\mathfrak{G}(U_1 \cap (K_1 - K_1))) \subset V$$
U_2 is relatively open in $K_1 - K_1$ and satisfies $\mathfrak{G}U_2 = U_2 \not\supset K_1 - K_1$ because:
  - $T \in \mathfrak{G}$ is invertible ($\mathfrak{G}$ is a group) hence T maps open sets to open sets and $T(A \cap B) = TA \cap TB$ for sets A, B.
  - Since T is linear, for any A
  $$T \operatorname{conv}(A) = \operatorname{conv}(TA) \tag{7}$$
  - Since $\mathfrak{G}$ is a group, $\mathfrak{G}\mathfrak{G}A = \mathfrak{G}A$.

- By continuity, $\mathfrak{G}U_2 = \overline{\mathfrak{G}U_2}$.
- Let $\delta = \inf\{a : a > 0, aU_2 \supset K_1 - K_1\} \geq 1$, by compactness $\delta < \infty$. Let $U := \delta U_2$.
- For any $\varepsilon \in (0,1)$, note
$$(1+\varepsilon)U \supset K_1 - K_1 \not\subset (1-\varepsilon)\overline{U}$$

Next, we use compactness of K_1 to get a center p which is within $(1-\varepsilon)\bar{U}$ of all of K_1. Combined with $(1-\varepsilon)\bar{U} \not\supset K_1 - K_1$, we can define $K_2 \subset\neq K_1$ with the desired properties.

- For the relatively open cover $\{2^{-1}U + k\}_{k \in K_1}$ of K_1, let $\{k_i\}_{i=1}^n$ index a finite subcover and define (the center) $p = \frac{1}{n}\sum_{i=1}^n k_i$. Then for all $k \in K_1$, $k_i - k \in 2^{-1}U$ for some $i \in [n]$ and since $k_j - k \in (1+\varepsilon)U$ for all $j \neq i$ we have
$$p \in \frac{1}{n}(2^{-1}U + (n-1)(1+\varepsilon)U) + k$$
Setting $\varepsilon = \frac{1}{4(n-1)}$ we get $p \in (1 - \frac{1}{4n})U + k$ for each $k \in K_1$, i.e. every point in K_1 is within $(1-\frac{1}{4n})U$ of the "center" p.
- Let
$$K_2 = K_1 \cap \bigcap_{k \in K_1} \left(\left(1 - \frac{1}{4n}\right)\bar{U} + k\right) \neq \emptyset$$
- Because $(1-\frac{1}{4n})\bar{U} \not\supset K_1 - K_1$, we have $K_2 \subsetneq K_1$.
- K_2 is closed and convex
- Further, since $T(a\bar{U}) \subset a\bar{U}$ for $T \in \mathfrak{G}$, we have
$$T(a\bar{U} + k) \subset a\bar{U} + Tk \qquad \text{forall } T \in \mathfrak{G}, k \in K_1$$
- Recalling $TK_1 \subset K_1$ for $T \in \mathfrak{G}$ and that $\mathfrak{G}$ is a group, we find that $TK_1 = K_1$ and $\mathfrak{G}K_2 \subset K_2$, contradicting minimality of K_1. a contradiction

□

2 Lecture 2: Measure theory

2020-01-23

Throughout our discussion, all topological spaces are assumed Hausdorff unless explicitly noted otherwise.

2.1 Construction of Haar measure

Definition 21

A *topological group* is a group equipped with a topology such that the group operations $(g,h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous.

Theorem 22 (*Existence of Haar Measure*)

{thm:haar-measure}

Let G be a compact topological group and $\mathcal{C}(G)$ the set of continuous maps $G \to \mathbb{R}$. Then there is a unique linear form $m : \mathcal{C}(G) \to \mathbb{R}$ such that

1. *$m(f) \geq 0$ for $f \geq 0$ (positive)*
2. *$m(\mathbb{1}) = 1$ (normalized)*
3. *$m({}_sf) = m(f)$ where ${}_sf(g) = f(s^{-1}g)$ for $s, g \in G$ (left invariant)*
4. *$m(f_s) = m(f)$ where $f_s(g) = f(gs)$ (right invariant)*

m is called the Haar measure *on G.*

We will need the following theorem to relate compactness with equicontinuity:

Theorem 23 (***Generalization of Arzela-Ascoli***)

{thm:arzela-ascoli} *Let X be a compact Hausdorff space. A subset of $\mathbb{R}$-valued continuous functions $F \subset \mathcal{C}(X)$ is relatively compact in topology induced by uniform norm $\|\cdot\|_\infty \Leftrightarrow F$ is equicontinuous and pointwise bounded.*

Proof of Theorem 22. Fix $f \in \mathcal{C}(G)$ and let $\mathcal{C}_f$ denote the convex hull of all left translates of f, i.e. $g \in \mathcal{C}_f$ are finite sums of form

$$g(x) = \sum_{\text{finite}} a_i f(s_i x), \qquad a_i > 0, \sum_{\text{finite}} a_i = 1, s_i \in G$$

Clearly $\|g\|_\infty \le \|f\|_\infty < \infty$, thus $\mathcal{C}_f(x) = \{g(x) : g \in \mathcal{C}_f\}$ is bounded for all $x \in G$ hence $\mathcal{C}_f$ is pointwise bounded.

As f is a continuous function on compact G, it is uniformly continuous hence for $\varepsilon > 0$ there exists a neighborhood V_ε of the identity $e \in G$ such that

$$y^{-1}x \in V_\varepsilon \Rightarrow |f(x) - f(y)| \le \varepsilon$$

Since $(s^{-1}y)^{-1}s^{-1}x = y^{-1}x$, we also have

$$y^{-1}x \in V_\varepsilon \Rightarrow |{}_sf(y) - {}_sf(x)| < \varepsilon$$

Since $g \in \mathcal{C}_f$ are convex combinations of ${}_sf$, by the triangle inequality

$$y^{-1}x \in V_\varepsilon \Rightarrow |g(y) - g(x)| < \varepsilon$$

As this works for any $g \in \mathcal{C}_f$, we have that $\mathcal{C}_f$ is equicontinuous.

By Theorem 23 (Generalization of Arzela-Ascoli), $\mathcal{C}_f$ is relatively compact in $\mathcal{C}(G)$, so its closure $K_f \coloneqq \overline{\mathcal{C}_f}$ is compact (and still convex).

Consider G acting on $\mathcal{C}(G)$ by left translation $f \mapsto {}_sf$. Notice $G\mathcal{C}_f \subset \mathcal{C}_f$ (as $\mathcal{C}_f$ already contains all finite convex combinations of all left translations of f) and hence $GK_f \subset K_f$ as well.

Furthermore, $\|{}_sf - {}_sg\|_\infty = \|f - g\|_\infty$ so G acts as a group of isometries on $\mathcal{C}(G)$. In particular, this group is equicontinuous (with the same $U = V$ in Definition 16).

Taking $\mathfrak{G} = G$ and $K = K_f$ in Theorem 20 (Kakutani Fix Point Theorem), there is a fixed point $g \in K_f$ of this action of G on K_f which satisfies

$${}_sg = g\ (\forall s \in G) \quad \Rightarrow \quad g(s^{-1}) = {}_sg(e) = g(e) = c\ (\forall s \in G)$$

for some constant $c \in \mathbb{R}$ (which we will later use to define $m(f) \coloneqq c$).

We first show there is only one constant function in K_f, so the fix point $Gg = \{g\} = \{c\mathbb{1}\}$ is unique and $m(f) = c$ is well defined. For any constant function $c\mathbb{1} \in K_f$ and $\varepsilon > 0$, we can (because $K_f = \overline{\mathcal{C}_f}$) find $\{s_1, \ldots, s_n\} \subset G$ and $a_i > 0$ such that

$$\sum_{i=1}^n a_i = 1, \quad \text{and} \quad \left| c - \sum_{i=1}^n a_i f(s_i x) \right| < \varepsilon \qquad (\forall x \in G) \tag{8}$$ {{eq:combo-close-to-constant}}

for any $\varepsilon > 0$.

Similarly, consider the same construction as before expect now use right translations of f (i.e. using the opposite group G' of G, or the function $f' = f(x^{-1})$, obtaining relatively compact set $\mathcal{C}'_f$ with compact convex closure K'_f with fix point $g' = c'\mathbb{1}$). Approximating $c'\mathbb{1}$ using $\mathcal{C}'_f$, we have

$$\left| c' - \sum_j b_j f(x t_j) \right| < \varepsilon \qquad (\text{for some } t_j \in G,\ b_j > 0 \text{ with } \sum_j b_j = 1) \tag{9}$$

Opposite group

The opposite group g' of the group G is the group that coincides with G as a set but has group operation $(x, y) \mapsto y^{-1}x^{-1}$

Summing over i

$$\left| c' - \sum_{i,j} a_i b_j f(s_i t_j) \right| < \varepsilon \sum_i a_i = \varepsilon$$

Operating symmetrically on Eq. (8) (multiply by b_i and put $x = t_i$) shows

$$\left| c - \sum_{i,j} a_i b_j f(s_i t_j) \right| < \varepsilon$$

Together, we have $|c' - c| < 2\varepsilon$ so taking $\varepsilon \to 0$ shows $c' = c$. Since $c\mathbb{1} \in K_f$ was an arbitrary constant function, we have that the constant function in K_f is actually unique and so the function $m(f) \coloneqq c \in K_f$ is well defined. Moreover, $m(f)\mathbb{1}$ is the *only* constant function which can be arbitrary well approximated by convex combinations of left or right translates of f.

The following properties are obvious:

- $m(\mathbb{1}) = 1$ since $K_f = \{1\}$ for $f = \mathbb{1}$
- $m(f) \geq 0$ if $f \geq 0$
- $m({}_s f) = m(f) = m(f_s)$ (since $K_{{}_s f} = K_f$, $K'_f = K'_{f_s}$, and uniqueness of $m(f)\mathbb{1}$ being the only constant function approximable by both K_f and K'_f)
- $m(af) = am(f)$ for any $a \in \mathbb{R}$ (since $K_{af} = K_f$)

To show m is linear, it suffices (due to the last bullet above) to show that m is additive. Fix $f, g \in \mathcal{C}(G)$. Approximate $m(f)$ using K_f to get

$$\left| m(f) - \sum_{\text{finite}} a_i f(s_i x) \right| \tag{10}$$

{{eq:approx-mf-using-Kf}}

Define $h(x) = \sum_{\text{finite}} a_i g(s_i x)$ using the same a_i and s_i and approximate $m(h)$ using $\mathcal{C}_h$ to get

$$\left| m(h) - \sum_{\text{finite}} b_j h(t_j x) \right| < \varepsilon$$

Since $h \in \mathcal{C}_g$, we have $\mathcal{C}_h \subset \mathcal{C}_g$ hence $K_h \subset K_g$. But $m(g)\mathbb{1} \in K_g$ is the only constant function so $m(h) = m(g)$ and (after expanding the definition of h) we have

$$\left| m(g) - \sum_{i,j<\infty} a_i b_j g(s_i t_j x) \right| < \varepsilon$$

On the other hand, multiplying Eq. (10) by b_j replacing x with $t_j x$, summing over j, and finally adding with the above ineqeuality gives

$$|m(f) + m(g) - \sum_{i,j} a_i b_j (f+g)(s_i t_j x)| < 2\varepsilon$$

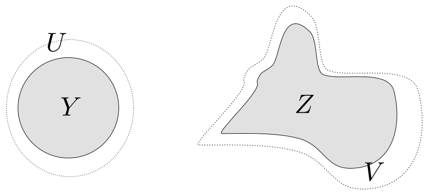
Thus $m(f) + m(g) \in K_{f+g}$, establishing additivity. Note that the only constant in K_{f+g} is $(m(f) + m(g))\mathbb{1}$. □

2.2 Facts from topology

We now want to head towards some integration against probability measures defined on spaces more abstract than $\mathbb{R}^n$.

Definition 24

A topological space X is *normal* if for any disjoint closed sets Y and Z there exists disjoint open sets U and V such that $Y \subset U$ and $Z \subset V$.



{fig:normal-topological-space} Figure 1: Normal topological spaces admit separating closed sets with two disjoint open sets

Definition 25

X is *completely regular* (*Tychonoff* if X is also Hausdorff) if for all $y \in X$ and every closed $Z \subset X \setminus \{y\}$ there exists $f : X \to [0,1]$ continuous such that $f(y) = 0$ and $f(z) = 1$ for all $z \in Z$. We say y and Z are separated by a (Urysohn) function.

Corollary 26 (*Urysohn's Lemma*)

Every normal space is completely regular.

Lemma 27

A compact (Hausdorff) space is normal hence completely regular.

Proof. Fix disjoint closed Y and Z and let $y \in Y$. Consider the open cover of Z given by $\{V_{y,z} : z \in Z\}$ where each $V_{y,z} \in N(z)$ is disjoint from some $U_{y,z} \in N(y)$ (existence ensured by Hausdorff). By compactness, there exists a finite subcover $\{V_{y,z_i}\}_{i=1}^n$. For each of these V_{y,z_i}, let $U_{y,z_i} \in N(y)$ denote the corresponding disjoint neighborhood of y and consider

$$U'_y = \bigcap_{i=1}^n U_{y,z_i} \in N(y)$$

U'_y is open because it is the intersection of finitely many open sets. It is also disjoint from

$$V'_y \coloneqq \bigcup_{i=1}^n V_{y,z_i}$$

which contains B and is also open.

Now consider the open cover $\{U'_y : y \in Y\}$, let $\{U'_{y_i}\}_{i=1}^n$ be a finite subcover, and let $U = \cup_{i=1}^n U'_{y_i}$. Analogously, let $V = \cap_{i=1}^n V'_{y_i}$ where V'_y is given above (open cover of B and disjoint from U'_y). Then $U \supset Y$ and $V \supset Z$ provide two disjoint separating open sets. □

Lemma 28

{lem:completely-regular-equals-initial-topo-cts} *A topological space (X, τ) is completely regular (i.e. Tychonoff) space iff the original topology coincides with the initial topology $\sigma(X, \mathcal{C}(X))$ i.e. the smallest topology that makes every function in $\mathcal{C}(X)$ continuous.*

Proof. We only show $\Rightarrow$. Let U be τ-open and for $x \in U$ pick an Urysohn function $f \in \mathcal{C}(X)$ such that $f(x) = 0$ and $f(U^c) = 1$. Then $V_x = \{y : f(y) < 1\} = f^{-1}((-\infty, 1))$ is a $\sigma(X, \mathcal{C}(X))$-open neighborhood of x contained in U, so $U = \cup_{x \in U} V_x$ is $\sigma(X, \mathcal{C}(X))$-open. Since $\sigma(X, \mathcal{C}(X))$ is minimal, we have $\tau = \sigma(X, \mathcal{C}(X))$. □

2.3 Radon, Borel, and Baire measures

Definition 29

A non-negative set function $m : 2^X \to [0, +\infty]$ on X is an *outer measure on X* (or Carathéodory outer measure) if:

1. $m(\emptyset) = 0$

2. $A \subset B \Rightarrow m(A) \leq m(B)$ (monotone)
3. $m\left(\cup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} m(A_n)$ for all $A_n \subset X$. (countable subadditivity)

Definition 30

Let $m : 2^X \to [0, +\infty]$ be a non-negative set function satisfying $m(\emptyset) = 0$. A set $A \subset X$ is *Carathéodory measurable wrt m* (Carathéodory m-measurable) if for any $E \subset X$

$$m(E) = m(E \cap A) + m(E \setminus A)$$

We use $\mathfrak{M}_m$ to denote the class of all Carathéodory m-measurable sets.

Theorem 31 (*Carthéodory construction*)

1. *$\mathfrak{M}_m$ is an algebra, m is additive on $\mathfrak{M}_m$*
2. *(Finite additivity) For all sequences of pairwise disjoint $A_i \in \mathfrak{M}_m$ and any $E \subset X$*

$$m\left(E \cap \bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} m(E \cap A_i)$$
$$m\left(E \cap \bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} m(E \cap A_i) + \lim_{n\to\infty} m\left(E \cap \bigcup_{i=n}^{\infty} A_i\right)$$

3. *If m is an outer measure on X, then $\mathfrak{M}_m$ is a σ-algebra, m is countably additive on $\mathfrak{M}_m$, and m is complete on $\mathfrak{M}_m$*

Remark 32. The outer measure is constructed such that it satisfies countable additivity on the measurable sets $\mathfrak{M}_m$.

Example 33

Let $\mathfrak{X}$ be a family of subsets of X such that $\emptyset \in \mathfrak{X}$. Given $\tau : \mathfrak{X} \to [0, +\infty]$ with $\tau(\emptyset) = 0$, set

$$m(A) = \inf\left\{\sum_{n=1}^{\infty} \tau(X_n) : X_n \in \mathfrak{X}, A \subset \cup_{n=1}^{\infty} X_n\right\}$$

where $\mathfrak{M}(A) = \infty$ in the absence of such sets X_n. Then m is an outer measure, denoted τ^*.

This is where the "outer" comes from: $\cup_n X_n \supset A$ is an outer approximation to A using (potentially overlapping) sets from $\mathfrak{X}$ hence $\sum_{n=1}^{\infty} \tau(X_n)$ is an overapproximation to the "size" of A. $m(A)$ is the best (i.e. smallest) overapproximation.

Recall the *Borel σ-algebra*, denoted $\mathcal{B}(X)$, is generated by all open sets.

Definition 34

The *Baire σ-algebra*, denoted by $\mathcal{B}a(X)$, is generated by sets of the form

$$\{x \in X : f(x) > 0\} \tag{11}$$ {{eq:functionally-open}}

where $f \in \mathcal{C}(X)$ (called *functionally open sets*).

Remark 35. $\mathcal{B}a(X)$ is the smallest σ-algebra where every $f \in \mathcal{C}(X)$ is measurable. It coincides (via a truncation and monotonicity argument) to the smallest one making every $f \in \mathcal{C}_b(X)$ measurable. Contrast this to Lemma 28, which shows that completely regular spaces are those with the smallest topology where every $f \in \mathcal{C}(X)$ is continuous.

Remark 36. Since the functionally open sets can be written as $f^{-1}((0, \infty))$ for continuous f, they are also Borel sets. Therefore, the class of Baire sets are contained in the class of Borel sets.

Lemma 37

{lem:metric-space-closed-set-variety}

In a metric space (X,d), any closed set S is the set of zeros of a continuous function (namely $d_S(x) = \inf_{s\in S} d(x,s)$). Hence, $\mathcal{B}(X) = \mathcal{B}a(X)$.

Lemma 38 (*Baire sets are countably determined*)

Every $A \in \mathcal{B}a(X)$ is determined by some countable family of functions, i.e. has the form

$$A = \{x : (f_i(x))_{i=1}^{\infty} \in B\} \quad \text{for some } f_i \in \mathcal{C}(X), B \in \mathcal{B}(\mathbb{R}^{\aleph_0})$$

Moreover, every set of this form is Baire and we can take $f_i \in \mathcal{C}_b(X)$.

Proof. We first show every set of the same form as A is Baire. True if B is closed, since Lemma 37 allows us to write $B = \phi^{-1}(0)$ for some continuous function $\phi : \mathbb{R}^{\aleph_0} \to \mathbb{R}$ so $\psi = x \mapsto \phi((f_n(x))_{n\geq 1})$ is continuous hence $A = \psi^{-1}(0)$ is also closed. But this is the converse.

For any fixed $\{f_n\}_{n\geq 1}$, the class of sets $B \in \mathcal{B}(\mathbb{R}^{\aleph_0})$ satisfying

$$\{x : (f_i(x))_{i\geq 1} \in B\} \in \mathcal{B}a(X)$$

is a σ-algebra containing $B = \prod_i(-\infty, a_i)$ where $a_i \neq \infty$ for only finitely many i. This is a basis for $\mathcal{B}(\mathbb{R}^{\aleph_0})$, thus $\mathcal{B}a(X)$ contains it and the two coincide (recall $\mathcal{B}a \subset \mathcal{B}$ since functionally determined sets are $\mathcal{B}$-open).

On the other hand??? ☐

A consequence of the monotone class theorem ???

Throughout, we consider (signed) measures of *bounded variation* unless explicitly denoted otherwise.

Definition 39

Let X be a topological space.

- A countably additive measure on $\mathcal{B}(X)$ is called a *Borel measure*
- A countably additive measure on $\mathcal{B}a(X)$ is called a *Baire measure*
- A Borel measure μ on X is called *Radon measure* if every $B \in \mathcal{B}(X)$ can be approximated from the inside by compact sets: for $\varepsilon > 0$ exists $K_\varepsilon \subset B$ such that $|\mu|(B \setminus K_\varepsilon) < \varepsilon$.

When are two Borel measures equal?

Lemma 40

If two Borel measures coincide on all open sets, thne they coincide on all Borel sets.

Proof. Split $\mu = \mu^+ - \mu^-$ and notice that each of the two components are nonnegative and coincide on open sets. By monotone class theorem, $\mu^+ = \mu^-$. ☐ finish

μ is Radon iff $|\mu|$ is Radon iff both μ^+ and μ^- are Radon.

Inner and outer approximation of measures on $\mathbb{R}^n$:

Theorem 41

$\mu \geq 0$ on $\mathcal{B}(\mathbb{R}^n)$, then any Borel set $B \subset \mathbb{R}^n$ and any $\varepsilon > 0$ exists U_ε open and F_ε closed such that $F_\varepsilon \subset A \subset U_\varepsilon$ and $\mu(U_\varepsilon \setminus F_\varepsilon) < \varepsilon$.

Proof. Let $\mathcal{A}$ the class of all sets $Ain\mathcal{B}$ such that $F_\varepsilon \subset A \subset U_e$ and $\mu(U_\varepsilon \setminus F_\varepsilon) < \varepsilon$ for some closed set F_ε and open set U_ε.

Every closed A is in $\mathcal{A}$, since we can take $F_\varepsilon = A$ and U_ε some open δ-nbd and consider $\delta \to 0$.

It suffices to show that $\mathcal{A}$ is a σ-algebra, since the closed sets generate $\mathcal{B}$. $\mathcal{A}$ is closed wrt complements, so it remains to verify closure under countable union.

Let $A_j \in \mathcal{A}$, $\varepsilon > 0$. Then exists closed F_j and open $U - j$ such that $F_j \subset A_j \subset U - j$ and $\mu(U_j \setminus F_j) < \varepsilon 2^{-j}$ for $j \in \mathbb{N}$.

The set $U = U_{j=1}^{\infty} U_j$ is open, and $Z_k = U_{j=1}^{k} F_j$ is closed.

Observe $Z_k \subset \cup_{j=1}^{\infty} A_j \subset U$ and for sufficiently large k $\mu(U \setminus Z_k) < \varepsilon$.

Indeed, $\mu(\cup_j^{\infty} \mu_j \setminus F_j) < 2$ ☐ ??

Definition 42

Set function $\mu \geq 0$ defined on $\mathcal{A} \subset 2^X$ is *tight* on $\mathcal{A}$ if $\forall \varepsilon > 0$ exists compact $K_\varepsilon \subset X$ such that $\mu(A) < \varepsilon$ for all $A \in \mathcal{A}$ that does not meet K_ε.

Additive set function μ of bounded variation on an algebra is *tight* if its total variation $|\mu|$ is tight.

A Borel measure is tight iff $\forall \varepsilon > 0$ exists compact K_ε such that $|\mu|(X \setminus K_\varepsilon) < \varepsilon$ (the "total variation measure").

The second definition is necessary to handle Baire sets.

Definition 43

μ is *regular* if $\forall A \in \mathcal{A}, \varepsilon > 0$, $\exists F_\varepsilon$ closed such that $F_\varepsilon \subset A$, $A \setminus F_\varepsilon \in \mathcal{A}$, and

??

Theorem 27 implies any Borel measure on $\mathbb{R}_n$ is regular, and the same proof works to show any Borel measure on metric space is regular.

Corollary 44 (*Baire measures are regular*)

{corr:baire-measure-regular}

Every Baire measure μ on topological space X is regular. Moreover, for every Baire set E and $\varepsilon > 0$, there exists a continuous function f on X such that $f^{-1}(0) \subset E$ and $|\mu|(E \setminus f^1(0)) < \varepsilon$.

3 Lecture 3: Measure theory

2020-01-28

Theorem 45 (*Extension to Radon measure*)

{thm:extend-tight-to-radon}

Suppose an algebra $\mathcal{A}$ of subsets of topological space X contains a base of the topology. Let μ be a regular additive set function of bounded variation on $\mathcal{A}$. If μ is tight, then it admits a unique extension to a Radon measure on X.

Proof. V.I. Bogachev, "Measure Theory" Theorem 7.3.2 ☐

Tightness is important because it says the whole space is inner approximable by a compact set.

Corollary 46

Let X be a completely regular spaace. Then every tight Baire measure μ on X admits a unique extensino to a Radon measure.

Proof. Every Baire measure is regular by Corollary 44, and since X is completely regular, functionally open sets form a base of the topology. Apply Theorem 49. ☐

This allows us to extend measures on the Baire σ-field to measures on the Borel σ-field.

Definition 47

A *vector lattice of functions* is a linear space of real functions on a nonempty set Ω such that $\max(f, g) \in \mathcal{F}$ for all $f, g \in \mathcal{F}$.

Notice $\min(f, g) = \max(-f, -g) \in \mathcal{F}$ and $|f| \in \mathcal{F}$.

Theorem 48 (*Daniell integration*)

{thm:daniell-integration}

Let $\mathcal{F}$ be a vector lattice of functions on a set Ω such that $\mathbb{1} \in \mathcal{F}$. Let L be a linear functional on $\mathcal{F}$ with:

- *$L(f) \geq f$ for all $f \geq 0$ (positive)*
- *$L(\mathbb{1}) = 1$*
- *$L(f_n) \to 0$ for every $f_n \downarrow 0$*

Then there exists a unique probability measure μ on $\mathcal{A} = \Sigma(\mathcal{F})$ generated by $\mathcal{F}$ such that $\mathcal{F} \subset \mathcal{L}^1(\boldsymbol{U})$ and

$$L(f) = \int_\Omega f d\mu, \qquad \forall f \in \mathcal{F}$$

Compare this with Riesz representation theorem

For X a compact space, L linear functional on $\mathcal{C}(X)$ with $L(\mathbb{1}) = 1$ and $L(f) \geq 0$ for $f \geq 0$ (positive linear functional), then $L(f) = \int_X f d\mu$ with unique regular Borel probability measure μ on X.
The relation is through Dini's theorem: If $\{f_n\} \subset \mathcal{C}(X)$, X compact, and $f_n(x) \downarrow 0$, then $\lim\limits_{n\to\infty} \sup_{x\in X} f_n(x) = 0$.

Proof. Denote $\mathcal{L}^+$ the set of all bounded functions f of the form $f(x) = \lim\limits_{n\to\infty} f_n(x)$, where $f_n \in \mathcal{F}$ are nonegative and increasing. $\{f_n\}$ is uniformly bounded, hence $\{L(f_n)\}$ is increasing and bounded by properties of L

Let $L(f) = \lim\limits_n L(f_n)$. We show that the extended functional is well-defined, coincides on bounded nonnegative functions in $\mathcal{F}$ with the original functional, and possesses the following properties:

1. $L(f) \leq L(g)$ for all $f, g \in \mathcal{L}^+$ with $f \leq g$
2. $L(f + g) = L(f) + L(g)$, $L(cf) = cL(f)$ for all $c \in [0, +\infty)$
3. $\min(f, g) \in \mathcal{L}^+$, $\max(f, g) \in \mathcal{L}^+$, and
$$L(f) + L(g) = L(\min(f, g)) + L(\max(f, g))$$
4. $\lim\limits_n f_n \in \mathcal{L}^+$ for every uniformly bounded increasing sequence of functions $f_n \in \mathcal{L}^+$, and $L(\lim\limits_n f_n) = \lim\limits_n L(f_n)$.

Let $\{f_n\}$ and $\{g_k\}$ be two increasing sequences of nonnegative functions in $\mathcal{F}$ with $\lim\limits_n f_n \leq \lim\limits_k g_k$. By linearity and positivity of L, $\lim\limits_n L(f_n) \leq \lim\limits_n L(g_k)$ proving (1).

From the hypotheses of the theorem if $\psi_m \uparrow \psi \in \mathcal{F}$ are all nonnegative functions then $L(\psi_m) \to L(\psi)$ again through linearity: $L(\psi) - L(\psi_m) = L(\psi - \psi_m) \to L(0) = 0$ proving (2).

To verify (4), suppose $f_{k,n} \uparrow f_n \in \mathcal{L}^+$. Let $g_m = \max_{n\leq m} f_{m,n}$, so $g_m \in \mathcal{F}$ increasing and $f_{m,n} \leq g_m \leq f_m$ for $n \leq m$. Therefore $\lim\limits_m f_m = \lim\limits_m g_m \in \mathcal{L}^+$ and by linearity for $n \leq m$

$$L(g_m) \leq L(g_{m+1}), \qquad L(f_{m,n}) \leq L(g_m) \leq L(f_m)$$

Hence $\lim\limits_m L(f_m) = \lim\limits_m L(g_m) = L(\lim\limits_m g_m) = L(\lim\limits_m f_m)$

$$\begin{array}{cccc}
g_1 & f_{1,1} & f_{2,1} & f_{1,3} \\
 & \rotatebox{-90}{$\leq$} & \rotatebox{-90}{$\leq$} & \rotatebox{-90}{$\leq$} \\
g_2 & f_{1,2} & f_{2,2} & f_{2,3} \\
 & \rotatebox{-90}{$\leq$} & \rotatebox{-90}{$\leq$} & \rotatebox{-90}{$\leq$} \\
 & \vdots & \vdots & \vdots \\
 & f_1 & f_2 & f_3
\end{array}$$

{fig:sketch-of-proof-of-4} Figure 2: sketch of proof of 4

Denote by $\mathcal{G}$ the class of all sets G with $\mathbb{1}_G \in \mathcal{L}^+$, and for $G \in \mathcal{G}$ define $\mu(G) = L(\mathbb{1}_G)$. Observe that min/max convert to indicators, so by (3) the class $\mathcal{G}$ is closed wrt finite intersection/union, and hence by countable union by (4).

Furthermore, μ is nonnegative monotone additive function on $\mathcal{G}$, with inclusion-exclusion, and $\mu(G_n) \uparrow \mu(G)$.

According to (TODO: Ref: thm 20) and closure of $\mathcal{G}$ under countable union, the function

$$\mu^*(A) = \inf\{\mu(G) : G \in \mathcal{G}, A \subset G\}$$

is countably additive measure on the class

$$\mathcal{B} = \{B \subset \Omega : \mu^*(B) + \mu^*(\Omega \setminus B) = 1\}$$

Uncertain about above theorem

Should check details of section 1.5 Borgachev

For (iii) we verify $\mathcal{A} = \sigma(\mathcal{F}) \subset \mathcal{B}$. If $f \in \mathcal{L}^+$, then $\{f > g\} \in \mathcal{G}$ for all c since

$$\mathbb{1}\{f > c\} = \lim_n \min(1, n \max(f - c, 0))$$

Hence $f \in \mathcal{L}^+$ are measurable wrt $\sigma(\mathcal{G})$, but they are also measurable wrt $\sigma(\mathcal{F})$ (since they are monotone limits of things in $\mathcal{F}$), so $\mathcal{G} \subset \sigma(\mathcal{L}^+) = \sigma(\mathcal{F})$ and by Dynkin π-λ we have $\sigma(\mathcal{G}) = \sigma(\mathcal{F}) = \mathcal{A}$. Thus, it suffices to show $\mathcal{G} \subset \mathcal{B}$.

For $G \in \mathcal{G}$, let $f_n \uparrow \mathbb{1}_G$ so

$$\mu^*(G) = \mu(G) = \lim_{n\to\infty} L(f_n)$$

and since $\mu^*(G) + \mu^*(\Omega \setminus G) \geq 1$, to show $G \in \mathcal{B}$ it suffices to prove $\mu^*(G) + \mu^*(\Omega \setminus G) \leq 1$ i.e.

$$\mu^*(\Omega \setminus G) \leq \lim_n L(\mathbb{1} - f_n)$$

But since $\mathbb{1} - f_n \downarrow \mathbb{1}_{\Omega\setminus G}$ and $U_c = \{1 - f_n > c\}$ contains $\Omega \setminus G$ hence belongs to $\mathcal{G}$, therefore

$$\mathbb{1}_{U_c} \leq c^{-1}(\mathbb{1} - f_n)$$
$$\mu^*(\Omega \setminus G) \leq \mu(U_c) = L(\mathbb{1}_{U_c}) \leq c^{-1} L(1 - f_n)$$

Take $c \to 1$ and $n \to \infty$.

It remains to prove $\mathcal{F} \subset \mathcal{L}^1(\mu)$ and that $L(f) = \int_\Omega f d\mu$. Approximate $f \in \mathcal{L}^+$ with $f \leq 1$ by simple functions

$$f_n = \sum_{j=1}^{2^n - 1} j 2^{-n} \mathbb{1}\{j 2^{-n} < x < (j+1) 2^{-n}\}$$
$$L(f_n) = \sum_{j=1}^{2^n - 1} j 2^{-n} \mu\{j 2^{-n} < x < (j+1) 2^{-n}\}$$

Hence

$$L(f_n) = \int_\Omega f_n d\mu$$

TODO: finish

The uniqueness of μ satisfying ?? follows from the fact that it is uniquely determined on the class $\mathcal{G}$, which is closed wrt finite intersections and generates $\mathcal{A}$. □

4 Lecture 3: Measure theory

2020-01-28

Theorem 49 (*Extension to Radon measure*)

{thm:extend-tight-to-radon} *Suppose an algebra $\mathcal{A}$ of subsets of topological space X contains a base of the topology. Let μ be a regular additive set function of bounded variation on $\mathcal{A}$. If μ is tight, then it admits a unique* *extension to a Radon measure on X.*

Proof. V.I. Bogachev, "Measure Theory" Theorem 7.3.2 ☐

Tightness is important because it says the whole space is inner approximable by a compact set.

Corollary 50

Let X be a completely regular spaace. Then every tight Baire measure μ on X admits a unique extensino to a Radon measure.

Proof. Every Baire measure is regular by Corollary 44, and since X is completely regular, functionally open sets form a base of the topology. Apply Theorem 49. ☐

This allows us to extend measures on the Baire σ-field to measures on the Borel σ-field.

Definition 51

A *vector lattice of functions* is a linear space of real functions on a nonempty set Ω such that $\max(f, g) \in \mathcal{F}$ for all $f, g \in \mathcal{F}$.

Notice $\min(f, g) = \max(-f, -g) \in \mathcal{F}$ and $|f| \in \mathcal{F}$.

Theorem 52 (*Daniell integration*)

{thm:daniell-integration}

Let $\mathcal{F}$ be a vector lattice of functions on a set Ω such that $\mathbb{1} \in \mathcal{F}$. Let L be a linear functional on $\mathcal{F}$ with:

- $L(f) \geq f$ *for all* $f \geq 0$ *(positive)*
- $L(\mathbb{1}) = 1$
- $L(f_n) \to 0$ *for every* $f_n \downarrow 0$

Then there exists a unique probability measure μ on $\mathcal{A} = \Sigma(\mathcal{F})$ generated by $\mathcal{F}$ such that $\mathcal{F} \subset \mathcal{L}^1(\boldsymbol{U})$ and

$$L(f) = \int_\Omega f d\mu, \qquad \forall f \in \mathcal{F}$$

Compare this with Riesz representation theorem

For X a compact space, L linear functional on $\mathcal{C}(X)$ with $L(\mathbb{1}) = 1$ and $L(f) \geq 0$ for $f \geq 0$ (positive linear functional), then $L(f) = \int_X f d\mu$ with unique regular Borel probability measure μ on X.
The relation is through Dini's theorem: If $\{f_n\} \subset \mathcal{C}(X)$, X compact, and $f_n(x) \downarrow 0$, then $\lim\limits_{n\to\infty} \sup_{x\in X} f_n(x) = 0$.

Proof. Denote $\mathcal{L}^+$ the set of all bounded functions f of the form $f(x) = \lim\limits_{n\to\infty} f_n(x)$, where $f_n \in \mathcal{F}$ are nonegative and increasing. $\{f_n\}$ is uniformly bounded, hence $\{L(f_n)\}$ is increasing and bounded by properties of L

Let $L(f) = \lim\limits_n L(f_n)$. We show that the extended functional is well-defined, coincides on bounded nonnegative functions in $\mathcal{F}$ with the original functional, and possesses the following properties:

1. $L(f) \leq L(g)$ for all $f, g \in \mathcal{L}^+$ with $f \leq g$
2. $L(f + g) = L(f) + L(g)$, $L(cf) = cL(f)$ for all $c \in [0, +\infty)$
3. $\min(f, g) \in \mathcal{L}^+$, $\max(f, g) \in \mathcal{L}^+$, and
$$L(f) + L(g) = L(\min(f, g)) + L(\max(f, g))$$
4. $\lim\limits_n f_n \in \mathcal{L}^+$ for every uniformly bounded increasing sequence of functions $f_n \in \mathcal{L}^+$, and $L(\lim\limits_n f_n) = \lim\limits_n L(f_n)$.

Let $\{f_n\}$ and $\{g_k\}$ be two increasing sequences of nonnegative functions in $\mathcal{F}$ with $\lim_n f_n \leq \lim_k g_k$. By linearity and positivity of L, $\lim_n L(f_n) \leq \lim_n L(g_k)$ proving (1).

From the hypotheses of the theorem if $\psi_m \uparrow \psi \in \mathcal{F}$ are all nonnegative functions then $L(\psi_m) \to L(\psi)$ again through linearity: $L(\psi) - L(\psi_m) = L(\psi - \psi_m) \to L(0) = 0$ proving (2).

To verify (4), suppose $f_{k,n} \uparrow f_n \in \mathcal{L}^+$. Let $g_m = \max_{n \leq m} f_{m,n}$, so $g_m \in \mathcal{F}$ increasing and $f_{m,n} \leq g_m \leq f_m$ for $n \leq m$. Therefore $\lim_m f_m = \lim_m g_m \in \mathcal{L}^+$ and by linearity for $n \leq m$

$$L(g_m) \leq L(g_{m+1}), \qquad L(f_{m,n}) \leq L(g_m) \leq L(f_m)$$

Hence $\lim_m L(f_m) = \lim_m L(g_m) = L(\lim_m g_m) = L(\lim_m f_m)$

$$\begin{array}{cccc}
g_1 & f_{1,1} & f_{2,1} & f_{1,3} \\
 & \rotatebox{90}{$\geq$} & \rotatebox{90}{$\geq$} & \rotatebox{90}{$\geq$} \\
g_2 & f_{1,2} & f_{2,2} & f_{2,3} \\
 & \rotatebox{90}{$\geq$} & \rotatebox{90}{$\geq$} & \rotatebox{90}{$\geq$} \\
 & \vdots & \vdots & \vdots \\
 & f_1 & f_2 & f_3
\end{array}$$

{fig:sketch-of-proof-of-4}

Figure 3: sketch of proof of 4

Denote by $\mathcal{G}$ the class of all sets G with $\mathbb{1}_G \in \mathcal{L}^+$, and for $G \in \mathcal{G}$ define $\mu(G) = L(\mathbb{1}_G)$. Observe that min/max convert to indicators, so by (3) the class $\mathcal{G}$ is closed wrt finite intersection/union, and hence by countable union by (4).

Furthermore, μ is nonnegative monotone additive function on $\mathcal{G}$, with inclusion-exclusion, and $\mu(G_n) \uparrow \mu(G)$.

According to (TODO: Ref: thm 20) and closure of $\mathcal{G}$ under countable union, the function

$$\mu^*(A) = \inf\{\mu(G) : G \in \mathcal{G}, A \subset G\}$$

is countably additive measure on the class

$$\mathcal{B} = \{B \subset \Omega : \mu^*(B) + \mu^*(\Omega \setminus B) = 1\}$$

Uncertain about above theorem

Should check details of section 1.5 Borgachev

For (iii) we verify $\mathcal{A} = \sigma(\mathcal{F}) \subset \mathcal{B}$. If $f \in \mathcal{L}^+$, then $\{f > g\} \in \mathcal{G}$ for all c since

$$\mathbb{1}\{f > c\} = \lim_n \min(1, n \max(f - c, 0))$$

Hence $f \in \mathcal{L}^+$ are measurable wrt $\sigma(\mathcal{G})$, but they are also measurable wrt $\sigma(\mathcal{F})$ (since they are monotone limits of things in $\mathcal{F}$), so $\mathcal{G} \subset \sigma(\mathcal{L}^+) = \sigma(\mathcal{F})$ and by Dynkin π-λ we have $\sigma(\mathcal{G}) = \sigma(\mathcal{F}) = \mathcal{A}$. Thus, it suffices to show $\mathcal{G} \subset \mathcal{B}$.

For $G \in \mathcal{G}$, let $f_n \uparrow \mathbb{1}_G$ so

$$\mu^*(G) = \mu(G) = \lim_{n \to \infty} L(f_n)$$

and since $\mu^*(G) + \mu^*(\Omega \setminus G) \geq 1$, to show $G \in \mathcal{B}$ it suffices to prove $\mu^*(G) + \mu^*(\Omega \setminus G) \leq 1$ i.e.

$$\mu^*(\Omega \setminus G) \leq \lim_n L(\mathbb{1} - f_n)$$

But since $\mathbb{1} - f_n \downarrow \mathbb{1}_{\Omega \setminus G}$ and $U_c = \{1 - f_n > c\}$ contains $\Omega \setminus G$ hence belongs to $\mathcal{G}$, therefore

$$\mathbb{1}_{U_c} \leq c^{-1}(\mathbb{1} - f_n)$$
$$\mu^*(\Omega \setminus G) \leq \mu(U_c) = L(\mathbb{1}_{U_c}) \leq c^{-1} L(1 - f_n)$$

Take $c \to 1$ and $n \to \infty$.

It remains to prove $\mathcal{F} \subset \mathcal{L}^1(\mu)$ and that $L(f) = \int_\Omega f d\mu$. Approximate $f \in \mathcal{L}^+$ with $f \leq 1$ by simple functions

$$f_n = \sum_{j=1}^{2^n - 1} j 2^{-n} \mathbb{1}\{j 2^{-n} < x < (j+1) 2^{-n}\}$$
$$L(f_n) = \sum_{j=1}^{2^n - 1} j 2^{-n} \mu\{j 2^{-n} < x < (j+1) 2^{-n}\}$$

Hence

$$L(f_n) = \int_\Omega f_n d\mu$$

TODO: finish

The uniqueness of μ satisfying ?? follows from the fact that it is uniquely determined on the class $\mathcal{G}$, which is closed wrt finite intersections and generates $\mathcal{A}$. ☐

5 Lecture 4: Daniell integration

2020-01-30

Exists vector lattice $\mathcal{F}$ containing constants i.e. $\mathbb{1} \in \mathcal{F}$. $\mathcal{L}^+$ consists of f such that $0 \leq f_n \uparrow f \leq c$ for $(f_n) \subset \mathcal{F}$. Let $\mathcal{G} = \{G \subset \Omega : \mathbb{1}_G \in \mathcal{L}^+\}$, $G \in \mathcal{G}$, $\mu(G) = L(\mathbb{1}_G) = \lim_n L(f_n)$, $\mathcal{A} := \sigma(\mathcal{F}) = \sigma(\mathcal{G})$.

Corollary 53

Suppose that in the previous theorem the class $\mathcal{F}$ is closed wrt uniform convergence. Let $\mathcal{G}_\mathcal{F}$ be the class of sets of the form $\{f > 0\}$, $f \geq 0$, $f \in \mathcal{F}$. Then $\mathcal{G}_\mathcal{F}$ generates $\mathcal{A} = \sigma(\mathcal{F})$ and we have

TODO

Proof. Suffices to verify $\mathcal{G}_\mathcal{F}$ equals the $\mathcal{F}$ introduced in theorem's proof. We previously showed $\{f > 0\} \in \mathcal{G}$ for all $f \geq 0$ in $\mathcal{F}$

On the other hand, if $G \in \mathcal{G}$ then $f_n \uparrow \mathbb{1}_G$ for some $f_n \geq 0$ in $\mathcal{F}$. Letting $f = \sum_{n=1}^\infty 2^{-n} f_n$, by uniform convergence of the series we have $f \in \mathcal{F}$. Clearly $f \geq 0$ and $G = \{f > 0\}$. ☐

A general fact about vector lattices where signed measures decompose into a positive part and negative part. If ν is a signed measure on Ω, then $\nu = \nu_+ - \nu_-$ for $\nu_\pm$ unique nonnegative meaasures with disjoint supports. Its total variation decomposes as:

$$\|\nu\| = \nu_+(\Omega) + \nu_-(\Omega)$$

Theorem 54

Let $\mathcal{F}$ be a vector lattice of bounded functions on a set Ω such that $\mathbb{1} \in \mathcal{F}$. Suppose that we are given a linear functional L on $\mathcal{F}$ that is continuous wrt $\|f\| = \sup_\Omega |f(x)|$, i.e.

$$\|L\| = \inf\{c : \|L(f)\| \leq c\|f\| \ \forall f \in \mathcal{F}\} < \infty$$

Then L can be represented as $L = L^+ - L^-$ where $L^+ \geq 0$, $L^- \geq 0$, and for all nonnegative $f \in \mathcal{F}$ we have

$$L^+(f) = \sup_{0 \leq g \leq f} L(g), \qquad L^-(f) = -\inf_{0 \leq g \leq f} L(g)$$

In addition, letting $|L| = L^+ + L^-$, we have for $f \geq 0$

$$|L|(f) = \sup_{0 \leq |g| \leq f} |L(g)|, \qquad \|L\| = L^+(1) + L^-(1)$$

Proof. Given two nonnegative $f, g \in \mathcal{F}$ and $h \in \mathcal{F}$ such that $0 \leq h \leq f + g$, can write $h = h_1 + h_2$ where $0 \leq h_1 \leq f$, $0 \leq h_2 \leq g$, $h_1, h_2 \in \mathcal{F}$. Just let $h_1 = \min(f, g)$ and $h_2 = h - h_1$ and verify.

Let L^+ be defined by the previous theorem. We first show additivity on nonnegative functions. For $f, g \in \mathcal{F}$ nonnegative, we have

$$L^+(f+g) = \sup\{L(h) : 0 \leq h \leq f+g\} = \sup\{L(h_1) + L(h_2) : 0 \leq h_1 \leq f, 0 \leq h_2 \leq g\} = L^+(f) + L^+(g)$$

where we used the previous decomposition.

Now we show additivity on arbitrary functions. Let $f = f_1 - f_2$, where f_1, f_2 non-negative. There might be multiple decompositions for the samae f, but still

$$L^+(f) = L^+(f_1) - L^+(f_2)$$

since $f_1 + f^- = f_2 + f^+$ and we showed L^+ is additive on nonnegative functions.

Define $L^- = L^+ - L$ and since $L^+(f) \geq L(f)$ for $f \geq 0$ we have that L^- is also nonnegative.

Finally,

$$\begin{aligned}\|L\| \leq \|L^+\| + \|L^-\| \\ &= L^+(1) + L^-(1) \\ &= 2L^+(1) - L^-(1) \\ &= \sup\{L(2\psi - 1) : 0 \leq \psi \leq 1\} \\ &\leq \sup\{L(h) : -1 \leq h \leq 1\} \\ &\leq \|L\|\end{aligned}$$

□

Corollary 55

{corr:lin-ftl-repr-meas} *Suppose in addition $L(f_n) \to 0$ for every $f_n \downarrow 0$. Then L^+ and L^- share this property as well, and are defined by nonnegative countably additive measures on $\sigma(\mathcal{F})$ and L has representation*

$$L(f) = \int_\Omega f d\mu, \qquad \forall f \in \mathcal{F}$$

with some signed countably additive measure μ on $\sigma(\mathcal{F})$.

Proof. TODO □

Here is an analogue of the Riesz representation theorem:

Theorem 56

Let X be a topological space. The formula

$$L(f) = \int_X f d\mu$$

establishes a one-to-one correspondence between Baire measures μ on X and continuous linear functionals L on $\mathcal{C}_b(X)$ with the property

$$\lim_n L(f_n) = 0$$

for every $f_n \downarrow f$.

Proof. Any measure $\boldsymbol{U}$ on $\mathcal{B}a(X)$ defines a continuous linear functional on $\mathcal{C}_b(X)$. Converse follows from Theorem 52 and Corollary 55. □

See "Banach limit"
Dini theorem.
We get a Riesz representation for compact spaces:

Theorem 57

K compact. Then every continuous linear functional L on $C(K)$ has a unique Radon measure μ such that

$$L(f) = \int_K f d\mu, \qquad \forall f \in \mathcal{C}(K)$$

Proof. TODO □

From now, we assume S to be locally compact, second countable, and Hausdorff (lcscH). Let $\mathcal{G}, \mathcal{F}, \mathcal{K}$ denote open, closed, and compact sets in S and put $\hat{\mathcal{G}} = \{G \in \mathcal{G}, \bar{G} \in \mathcal{K}\}$. Let $\hat{C}_+ = \hat{C}_+(S)$ denote the class of continuous functions $f : S \to \mathbb{R}_+$ with compact support (i.e. closure of the set $\{x \in S; f(x) > 0\}$).

We want to extend the idea of invariant (Haar) measure from just groups to more general spaces such as the sphere.

Theorem 58 (*Riesz representation*)

{thm:riesz-extend-fts-to-meas} *If S is lcscH, then every positive linear functional μ on $\hat{C}_+(S)$ extends uniquely to a measure on S that assigns finite mass to compact sets.*

Proof. Kallenberg, "Foundations of modern probability" □

Theorem 59

On every lcscH group G there exists, uniquely up to normalization, a left-invariant measure $\lambda \neq 0$ that assigns finite mass to compact sets. If G is compact, then λ is also right-invariant.

Proof. Kallenberg, "Foundations of modern probability" □

Definition 60

Given group G and space S, a *left action* of G on S is a mapping $(g, s) \mapsto gs$ such that $es = s$ and $(gh)s = g(hs)$ for any $g, h \in G$ and $s \in S$, where e denotes the identity element in G.

Similarly, a *right action* is a mapping $(s, g) \mapsto sg$ satisfying similar compatibility conditions.

The action is *transitive* if for all $s, t \in S$ there exists $g \in G$ such that $gs = t$ or $sg = t$ respectively.

All actions are assumed left henceforth.

When G is a topological group, we assume the action is a continuous $G \times S \to S$ map.

Definition 61

$h : G \to S$ is *proper* if $h^{-1}K$ is compact in G for any compact $K \subset S$.

If this holds for all $\pi_s(x) = xs$, $s \in S$, we say the group action is proper.

Definition 62

A memasure μ on S is G-invariant if $\mu(xB) = \mu B$ for any $x \in G$ and $B \in \mathcal{S}$. This is clearly equivalent to

$$\int f(xs)\mu(ds) = \mu f$$

for any measurable $f : S \to \mathbb{R}_+$ and $x \in G$.

Theorem 63

If we have lcshH group G acting transively and properly on lcscH space S. Then there exists a unique (up to normalization) G-invariant measure $\mu \neq 0$ on S which assigns finite mass to compact sets.

Proof. We first show existence. Fix $p \in S$ and let $\pi = x \mapsto xp : G \to S$. Letting λ be a left Haar measure on G, define the pushforward $\mu = \lambda \circ \pi^{-1}$ on S. Since π is proper and the Haar measure on G assigns finite mass to compact sets, μ is a measure on S that assigns finite mass to compact sets. To see G-invariance, for $f \in \hat{C}_+(S)$ and $x \in G$

$$\int_S f(xs)\mu(ds) = \int_G f(xyp)\lambda(dy) = \int_G f(yp)\lambda(dy) = \mu f$$

by invariance of λ.

Now we consider uniqueness. Let μ be abitrary G-invariant measure on S assigning finite mass to compact sets. Define the subgroup

$$K = \{x \in G : xp = p\} = \pi^{-1}\{p\}$$

(the stabilizer of p, subgroup leaving p fixed) and note K is compact (since π is proper). Let ν be the normalized Haar measure on K, and define

$$\bar{f}(x) = \int_K f(xk)\nu(dk), \quad x \in G, f \in \hat{C}_+(G)$$

At each point x, $\bar{f}$ takes f and "smooths things out" using K translated to x.

If $xp = yp$ then $y^{-1}xp = p$ and so $y^{-1}x =: h \in K$ which implies $x = yh$. Hence, left invariance of ν yields

$$\bar{f}(x) = \bar{f}(yh) = \int_K f(yhk)\nu(dk) = \int_K f(yk)\nu(dk) = \bar{f}(y)$$

So the mapping $f \mapsto f^*$ given by

$$f^*(s) = \bar{f}(\pi^1\{s\}) \equiv \bar{f}(x), \quad s = xp \in S, x \in G, f \in \hat{C}_+(G)$$

is well defined, and for any $B \subset (0, \infty)$ we have

$$(f^*)^{-1}B = \pi(\bar{f}^{-1}B) \subset \pi[(\operatorname{supp} f) \cdot K]$$

where $(\operatorname{supp} f) \cdot K$ is the support of f "convolved with K" via the group action. Hence, the RHS is compact (both $\operatorname{supp} f$ and K compact) and since π and the action are continuous. Therefore f^* has compact support.

Also, $\bar{f}$ is continuous (by group operation cts and dominated convergence), so $\bar{f}^{-1}[t, \infty)$ is closed and hence compact for every $t > 0$.

??? So f^* is something we can integrate against μ.

We may now define functional λ on $\hat{C}_+(G)$ by $\lambda f = \mu f^*$ for $f \in \hat{C}_+(G)$. Linearity and positivity of λ are clear from the corresponding properties of the mapping $f \mapsto f^*$ and the measure μ. We note that λ is finite on $\hat{C}_+(G)$ since μ is locally finite, so by Theorem 58 we can extend λ to a measure on G that assigns fintie mass to compact sets.

To see λ left invariant, for $f \in \hat{C}_+(G)$ and define $f_y(x) = f(yx)$. THen for $s = xp \in S$ and $y \in G$ we have

$$f_y^*(s) = \bar{f}_y(x) = \int_K \bar{f}_y(xk)\nu(dk) = \bar{f}(yx) = f^*(ys)$$

Hence by invariance of μ we have

$$\int_G f(yx)\lambda(dx) = \lambda f_y = \mu f_y^* = \int_S f^*(ys)\mu(ds) = \mu f^* = \lambda f$$

So λ is the Haar measure.

Now fix $g \in \hat{C}_+(S)$ and put $f(x) = g(xp) = g \circ \pi(x)$ for $x \in G$. THen $f \in \hat{C}_+(G)$ because $\{f > 0\} \subset \pi^{-1} \operatorname{supp} g$ which is compact since π is proper. By definitino of K, for $s = xp \in S$ we have

$$f^*(s) = \bar{f}(x) = \int_K f(xk)\nu(dk) = \int_K g(xkp)\nu(dk) = \int_K g(xp)\nu(dk) = g(s)$$

so we've found an inverse for the $*$ operation, so

$$\mu g = \mu f^* = \lambda f = \lambda(g \circ \pi) = (\lambda \circ \pi^{-1})g$$

which shows $\mu = \lambda \circ \pi^{-1}$. Since λ is unique up to normalization, the same thing is true for μ. $\square$

References

{blackwell1973} [BM73] David Blackwell and James B. MacQueen. "Ferguson Distributions Via Polya Urn Schemes". In: *Ann. Statist.* 1.2 (Mar. 1973), pp. 353–355. DOI: 10.1214/aos/1176342372. URL: https://doi.org/10.1214/aos/1176342372.

{ferguson1973} [Fer73] Thomas S. Ferguson. "A Bayesian Analysis of Some Nonparametric Problems". In: *Ann. Statist.* 1.2 (Mar. 1973), pp. 209–230. DOI: 10.1214/aos/1176342360. URL: https://doi.org/10.1214/aos/1176342360.